COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 946

THE GENERALIZED BASIS REDUCTION ALGORITHM

by László Lovász and Herbert E. Scarf

June 1990

THE GENERALIZED BASIS REDUCTION ALGORITHM

by

László Lovász and Herbert E. Scarf*

I. Introduction

Let C be a compact convex body in R^n, of positive volume and symmetric about the origin, and let L be the lattice of integer vectors in R^n. The body can be used to define a distance function $F(x) = \inf\{\lambda \geq 0 | x/\lambda \in C\}$, with the properties:

1. F(x) is convex,
2. $F(-x) = F(x)$,
3. $F(tx) = tF(x)$ for $t > 0$.

The dual body C^* is defined to be $\{y \,| y.x \leq 1 \text{ for all } x \in C\}$, and the dual distance function is $F^*(y) = \max_{x \in C} y.x$.

---

*The authors are respectively at the Departments of Computer Science, Eötvös Loránd University, Budapest and Princeton University, Princeton; and The Cowles Foundation, Yale University, New Haven. The research of the first author was supported by the Hungarian National Foundation for Scientific Research, and of the second author by the Program on Discrete Mathematics at the Cowles Foundation and NSF grant SES-8807167.

We wish to thank David Applegate, Imre Bárány, Judith Bender, Bill Cook, Mark Hartman, Ravi Kannan, Jeffrey Lagarias, Tom Rutherford, David Shallcross and Mike Todd for a number of stimulating conversations on the subject of this paper.

In order to determine a smallest non-zero lattice point according to the distance function F, we introduce the concept of a reduced basis with respect to F. Let $b^1, b^2, \ldots, b^n$ be a basis for the integer lattice L. For each i we project C, along the vectors $b^1, \ldots, b^{i-1}$, into the affine space $E_i = \langle b^i, \ldots, b^n \rangle$ obtaining C_i. In other words, $x = x_i b^i + \ldots + x_n b^n \in C_i$ if and only if there are $\alpha_1, \ldots, \alpha_{i-1}$ such that $x + \alpha_1 b^1 + \ldots + \alpha_{i-1} b^{i-1} \in C$. The lattice L_i, obtained by projecting L along $b^1, \ldots, b^{i-1}$ into $\langle b^i, \ldots, b^n \rangle$, is the set of integral linear combinations of the vectors $b^i, \ldots, b^n$.

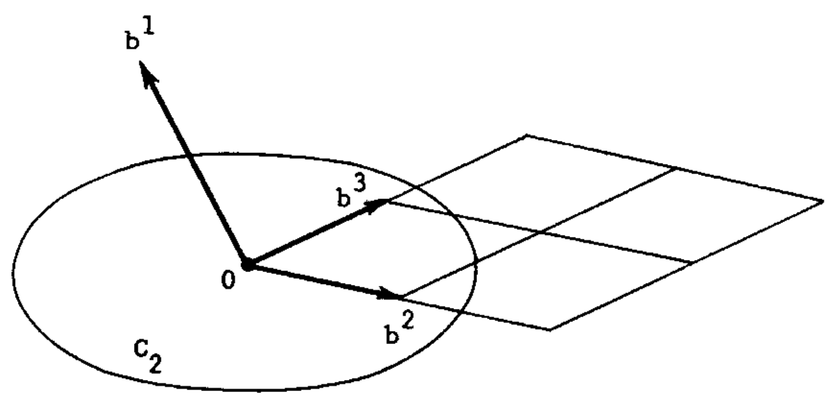

FIGURE 1

The distance function $F_i(x)$, associated with the projected body C_i, is defined for $x \in E_i$ by

$$F_i(x) = \min F(x + \alpha_1 b^1 + \ldots + \alpha_{i-1} b^{i-1}),$$

with the minimum taken over $\alpha_1, \ldots, \alpha_{i-1}$. The function may, of course, be defined for all x in R^n by the same formula; if $x = \Sigma x_j b^j$, then $F_i(x)$ will be independent of $x_1, \ldots x_{i-1}$.

Fix $0 < \varepsilon < 1/2$. The basis is reduced, for this ε, if the following two conditions hold for $i = 1,\ldots,n-1$:

1. $F_i(b^{i+1}+\mu b^i) \geq F_i(b^{i+1})$ for integral μ, and
2. $F_i(b^{i+1}) \geq (1-\varepsilon)F_i(b^i)$.

If C is an ellipsoid - or, alternatively, if C is the unit ball and the lattice is a general lattice in R^n - this definition of a reduced basis is identical with the definition in A. K. Lenstra, H. W. Lenstra, Jr. and L. Lovász(1982).

In Section II, we discuss the properties of a reduced basis, demonstrating, in particular, that for such a basis b^1 is an approximation to the shortest non-zero lattice point. In addition, b^i is an approximation to a lattice point realizing the ith successive minimum, according to Minkowski. We also provide a polynomial algorithm for fixed n which finds the shortest non-zero lattice point rather than an approximation.

In Section III, the basis reduction algorithm is described and shown to execute in polynomial time, for fixed n. In Section IV, we examine a special basis - the Korkine-Zolotarev basis - associated with a distance function F. Using the Korkine-Zolotarev basis, we provide an alternative demonstration of a theorem to be found in Kannan and Lovász (1988), that a lattice-free body K, in R^n, has associated with it a non-zero lattice point h, such that the width of the body in the direction h satisfies

$$\max_{x \in K}\{h.x\} - \min_{x \in K}\{h.x\} \leq c_0 n(n+1)/2,$$

with c_0, a universal constant.

Lenstra's polynomial algorithm (H. W. Lenstra, Jr.(1983)) for integer programming with a fixed number of variables makes use of the spherical basis reduction algorithm. He begins by a preliminary reduction to the problem of determining a lattice point in a convex polyhedron K, in R^n, defined by a system of linear inequalities $Ax \le c$. To find such a lattice point, the polyhedron is approximated by a ellipsoid E, and a hyperplane with integer normals h is found so that the width of the ellipsoid in the direction h,

$$\max_{x \epsilon E}(h.x) - \min_{x \epsilon E}(h.x)$$

is as small as possible, aside from a factor depending only on the number of variables, n. If this width is sufficiently large, the polyhedron is sure to contain a lattice point. In the alternative case, in which the width is not large, we consider the intersections of the polyhedron with the hyperplanes $hx = h_0$, with h_0 assuming all integral values between $\min_{x \epsilon E} h.x$ and $\max_{x \epsilon E} h.x$. The n dimensional problem is thereby reduced to the problem of determining a lattice point in one of a small number of n-1 dimensional polyhedra. Each of these polyhedra is then approximated by its own ellipsoid and the algorithm continues.

A non-zero lattice point h, which minimizes the width of the ellipsoid E, is a shortest non-zero lattice point for the body $(E\text{-}E)^*$, itself an ellipsoid. If this latter ellipsoid is transformed to a sphere by a linear transformation, an approximation to the shortest non-zero lattice point can be found using the spherical basis reduction algorithm for a general lattice.

The arguments of this note can be used to find a short non-zero lattice point for the body $(K - K)^*$ directly, thereby avoiding the series of ellipsoidal approximations. The basis reduction algorithm is applied to $C = (K - K)^*$, where $K = \{x | Ax \le c\}$, with the distance functions

$$F_i(\xi) = \min_\alpha F_1(\xi + \alpha_1 b^1 + \ldots + \alpha_{i-1} b^{i-1}),$$

$$\min_\alpha \max\{(\xi + \alpha_1 b^1 + \ldots + \alpha_{i-1} b^{i-1}).(x-y) | Ax \le c,\ Ay \le c\}$$

$$= \min_{\alpha,t,u} c.(t + u), \text{ subject to } t,u \ge 0,$$

$$tA = \xi + \alpha_1 b^1 + \ldots + \alpha_{i-1} b^{i-1},$$

$$uA = -(\xi + \alpha_1 b^1 + \ldots + \alpha_{i-1} b^{i-1}),$$ (from the duality theorem for linear programming.)

$$= \min_{\alpha,t,u} c.(t + u), \text{ subject to } t,u \ge 0,$$

$$tA - \alpha_1 b^1 - \ldots - \alpha_{i-1} b^{i-1} = \xi,$$

$$uA + \alpha_1 b^1 + \ldots + \alpha_{i-1} b^{i-1} = -\xi,$$

$$= \max \xi.(x - y), \text{ subject to}$$

$$Ax \le c,\ Ay \le c,\ b^1.(x - y) = 0, \ldots, b^{i-1}.(x - y) = 0$$ (using the duality theorem again.).

The general basis reduction algorithm requires the solution of many linear programs, and there are tradeoffs between using an ellipsoidal approximation to K, or working directly with the body, itself, to resolve the question of whether K contains a lattice point. A number of computational experiments are currently being attempted on integer programming problems of moderately large size to evaluate the merits of the two procedures.

II. Properties of a Reduced Basis

Theorem 1. Let $b^1, \ldots, b^n$ be a reduced basis. Then

$$F_{i+1}(b^{i+1}) \ge (1/2-\varepsilon) F_i(b^i) \text{ for } i = 1, \ldots, n-1.$$

Proof: We have the identity

$$\min F_i(x+\alpha b^i) = F_{i+1}(x)$$

with the minimum taken over all real α. Since we can round α to the nearest integer μ, it follows that

$$\min F_i(x+\mu b^i) \le F_{i+1}(x) + 1/2\ F_i(b^i), \tag{1}$$

with the minimum taken over integer μ. If x is taken to be b^{i+1} then (1), in conjunction with the definition of a reduced basis, tells us that

$$(1-\varepsilon)F_i(b^i) \le F_i(b^{i+1}) = \min F_i(b^{i+1}+\mu b^i)$$
$$\le F_{i+1}(b^{i+1}) + 1/2\ F_i(b^i). \otimes$$

Theorem 2. Let $b^1,\ldots,b^n$ be a reduced basis, and let

$\lambda_1 = \min F(h)$, for all non-zero lattice points h.

Then $\lambda_1 \ge F(b^1).(1/2-\varepsilon)^{n-1}$.

Proof: Let $h = l_1 b^1 + \ldots + l_k b^k$, with $l_1, \ldots, l_k$ integral and l_k different from zero, be a shortest non-zero lattice point according to the distance function F. Then

$$\lambda_1 = F(h) \ge F_k(h) = |l_k| F_k(b^k) \ge F(b^1).(1/2-\varepsilon)^{k-1}. \otimes$$

Theorem 2 states that the first vector, b^1, in a reduced basis is an approximation to the shortest non-zero lattice point. In a similar fashion the other basis vectors approximate the <u>successive minima</u> of the lattice with respect to the distance function.

Definition: $\lambda_1, \ldots, \lambda_n$ are the <u>successive minima</u> of the lattice with respect to F if there are lattice points $h^1, \ldots, h^n$, with $\lambda_i = F(h^i)$, such that for each $i = 1,\ldots,n$, h^i is the shortest lattice point which is linearly independent of $h^1, \ldots, h^{i-1}$.

An equivalent definition is $\lambda_i = \min \{\lambda | F(x) \le \lambda$ contains i linearly independent lattice points$\}$. The successive minima λ_i are uniquely defined by the distance function F, but there may be more than one set of lattice points h^i which realize these values. We have the following generalization of Theorem 2.

Theorem 3. Let $b^1,\ldots,b^n$ be a reduced basis.

Then for $i = 1, \ldots, n$,

$$F_i(b^i)(1/2-\varepsilon)^{n-i} \le \lambda_i \le F_i(b^i)/(1/2-\varepsilon)^{i-1}.$$

Proof: We begin by constructing a basis $c^1,\ldots,c^n$ for the lattice with

(2) $$F_1(c^i) \le F_i(b^i)/(1/2-\varepsilon)^{i-1},$$

thereby demonstrating the right hand side of the inequality in Theorem 3.

Again we use the inequality

$$\min F_i(x+\mu b^i) \le F_{i+1}(x) + 1/2\ F_i(b^i),$$

with the minimum taken over integer μ. If x is taken to be b^{i+1} then this inequality implies

$$F_i(b^{i+1}+\mu_{i+1,i}b^i) \le F_{i+1}(b^{i+1}) + 1/2\ F_i(b^i)$$

for some integral $\mu_{i+1,i}$. Apply the inequality again with i+1 replaced by i and $x = b^{i+1}+\mu_{i+1,i}b^i$, obtaining

$$F_{i-1}(b^{i+1}+\mu_{i+1,i}b^i+\mu_{i+1,i-1}b^{i-1}) \le$$
$$F_{i+1}(b^{i+1}) + 1/2\ F_i(b^i) + 1/2\ F_{i-1}(b^{i-1})$$

for some integral $\mu_{i+1,i}$ and $\mu_{i+1,i-1}$. Continuing, we see that

(3) $$F_1(b^{i+1} + \Sigma_{j=1}^{i}\mu_{i+1,j}b^j) \le F_{i+1}(b^{i+1}) + (1/2)\Sigma_1^i F_j(b^j)$$

for some integral $\mu_{i+1,j}$, $j=1,\ldots,i$.

We use this construction to define

(4) $$c^{i+1} = b^{i+1} + \Sigma_{j=1}^{i}\mu_{i+1,j}b^j.$$

Estimating $\Sigma F_j(b^j)$ by means of Theorem 1, we see that

$$F_1(c^{i+1}) \le F_{i+1}(b^{i+1}).\{1 + 1/2\ \Sigma 1/(1/2-\varepsilon)^{i+1-j}\}$$
$$\le\ F_{i+1}(b^{i+1})/(1/2-\varepsilon)^i.$$

To demonstrate the inequalities on the left hand side of Theorem 3, we write

$$h^1 = l_{11}b^1 + l_{12}b^2 + \ldots + l_{1n}b^n$$
$$\vdots$$
$$h^i = l_{i1}b^1 + l_{i2}b^2 + \ldots + l_{in}b^n$$
$$\vdots$$
$$h^n = l_{n1}b^1 + l_{n2}b^2 + \ldots + l_{nn}b^n,$$

with l_{ij} integral and with h^i linearly independent lattice points which realize the successive minima, i.e. $F_1(h^i) = \lambda_i$.

For each index i, there must be a pair of indices j and k with $j \le i \le k$ such that $l_{jk} \ne 0$, since otherwise

$$h^1 = l_{11}b^1 + \ldots + l_{1,i-1}b^{i-1}$$
$$\vdots$$
$$h^i = l_{i1}b^1 + \ldots + l_{i,i-1}b^{i-1},$$

and the vectors $h^1, \ldots, h^i$ would be linearly dependent. For each i, therefore, let k be the largest index such that $l_{jk} \ne 0$ for some $j \le i \le k$. But then, since $|l_{jk}| \ge 1$,

$$\lambda_i \ge \lambda_j = F_1(h^j) \ge F_k(h^j) = |l_{jk}|.F_k(b^k)$$
$$\ge F_i(b^i).(1/2-\varepsilon)^{k-i}$$
$$\ge F_i(b^i).(1/2-\varepsilon)^{n-i}.$$

This demonstrates Theorem 3. ⊗

According to Minkowski, the successive minima satisfy the inequality

$$\lambda_1\ldots\lambda_n.\mathrm{vol}(C) \le 2^n.$$

We can show that the basis $c^1,\ldots,c^n$, defined by (4), approximates this result in the following sense:

Theorem 4. Let $b^1,\dots,b^n$ be a reduced basis with respect to F. Then the basis $c^1,\dots,c^n$ satisfies

$$F_1(c^1).F_1(c^2)\dots F_1(c^n).\mathrm{vol}(C) \le 2^n/(1/2 - \varepsilon)^{n(n-1)/2}.$$

Proof: The proof depends on the fact that for any basis

$$F_1(b^1).F_2(b^2)\dots F_n(b^n).\mathrm{vol}(C) \le 2^n. \tag{5}$$

To demonstrate (5), let us assume, by induction, that this inequality is satisfied for the n-1 dimensional body C_2 obtained by projecting b^1 into the affine space $\langle b^2,\dots,b^n\rangle$, so that

$$F_2(b^2)\dots F_n(b^n).\mathrm{vol}(C_2) \le 2^{n-1}.$$

But then $\mathrm{vol}(C) = \int_{C_2} l(x)dx$, with $l(x)$ the length of the intersection of the line $x + \alpha b^1$ with C. From the symmetry and convexity of C, $l(x) \le l(0) = 2/F_1(b^1)$ so that $\mathrm{vol}(C) \le 2\mathrm{vol}(C_2)/F_1(b^1)$, thereby demonstrating (5). Theorem 4 follows from the previously established inequality

$$F_1(c^i) \le F_i(b^i)/(1/2 - \varepsilon)^{i-1}. \otimes$$

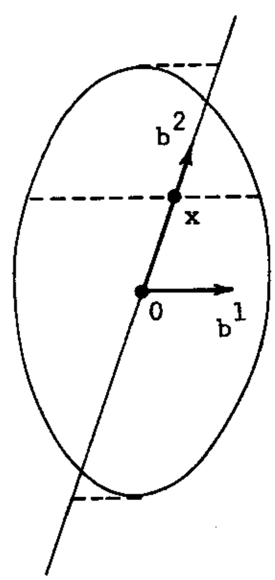

FIGURE 2

If we are given a reduced basis, then a shortest vector h^1 can be calculated in polynomial time for fixed n. We do this by establishing bounds on the coordinates of lattice points satisfying $F_1(h) \le F_1(b^1)$. Let $h = \Sigma l_j b^j$ be such a vector. Then $F_1(b^1) \ge F_1(h) \ge F_n(h) = |l_n| . F_n(b^n)$ so that $|l_n| \le F_1(b^1)/F_n(b^n) \le 1/(1/2-\varepsilon)^{n-1}$.

Now let us suppose that the coordinates $l_n, \ldots, l_{i+1}$ have been selected. We find bounds for l_i as follows: Find the real α which minimizes $F_i(l_n b^n + \ldots + l_{i+1} b^{i+1} + \alpha b^i)$. If the minimum is greater than $F_1(b^1)$ then there is no h with these final n-i coordinates satisfying $F_1(h) \le F_1(b^1)$. If, on the other hand, the minimum is less than or equal to $F_1(b^1)$ then since $F_i(l_n b^n + \ldots + l_{i+1} b^{i+1} + l_i b^i) = F_i(h) \le F_1(h) \le F_1(b^1)$, and F_i is symmetric and convex, we obtain

$$|l_i - \alpha| . F_i(b^i) \le 2F_1(b^1) \text{ or}$$
$$|l_i - \alpha| \le 2/(1/2-\varepsilon)^{i-1}.$$

This provides us with a tree of depth n and with a "small" number of branches at each node in which to search for the coordinates of the shortest vector. If the tree is used to calculate the shortest non-zero lattice point in actual numerical examples, the estimate $(1/2-\varepsilon)^{i-1}$ should be replaced by $F_i(b^i)/F_1(b^1)$, which may be considerably smaller.

If we look for the ith successive minimum by considering those h with $F_1(h) \le F_i(b^i)/(1/2-\varepsilon)^{i-1}$ we obtain precisely the same set of inequalities for $l_n, \ldots, l_i$, but we do not have similar bounds for the first i-1 coordinates of h. This yields a "small" number of hyperplanes of dimension i-1, one of which contains a lattice point which realizes the ith successive minimum.

III. The Basis Reduction Algorithm

An algorithm for finding a reduced basis may easily be described. We begin with an initial basis $a^1, a^2, \ldots, a^n$ for the lattice, and move through a sequence of bases $b^1, b^2, \ldots, b^n$ according to the following rules: At each step of the algorithm, we consider the first index i for which one of the conditions

1. $F_i(b^{i+1}+\mu b^i) \geq F_i(b^{i+1})$ for integral μ, and
2. $F_i(b^{i+1}) \geq (1-\varepsilon)F_i(b^i)$.

is not satisfied.

If the first condition is not satisfied, we replace b^{i+1} by $b^{i+1}+\mu b^i$, with μ the integer which minimizes $F_i(b^{i+1}+\mu b^i)$. If, after this replacement, the second condition obtains, we move to level i+1. If the second condition is not satisfied, we interchange b^i and b^{i+1} and move to the preceding level i-1, unless i = 1, in which case we remain at level 1.

In order to demonstrate convergence of the algorithm we consider the vector

$$F_1(b^1), \ldots, F_i(b^i), F_{i+1}(b^{i+1}), \ldots, F_n(b^n),$$

and remark that the maximum value of the components of the vector does not increase at any step of the basis reduction algorithm. If we replace b^{i+1} by $b^{i+1}+\mu b^i$, none of the terms change; if b^i and b^{i+1} are interchanged, $F_i(b^i)$ becomes $F_i(b^{i+1}) \leq (1-\varepsilon)F_i(b^i)$ and $F_{i+1}(b^{i+1})$ is replaced by

$$\begin{aligned} &\min F(b^i + \alpha_1 b^1 + \ldots + \alpha_{i-1} b^{i-1} + \alpha_{i+1} b^{i+1}) \\ \leq\ &\min F(b^i + \alpha_1 b^1 + \ldots + \alpha_{i-1} b^{i-1}) = F_i(b^i). \end{aligned}$$

It follows that at any step in the algorithm, $\max F_i(b^i) \leq \max F_i(a^i)$ equal to, say, U.

The basis reduction algorithm is known to converge in polynomial time, including the number of variables, n, for $F(x) = |x|$, and a general lattice given by an integer basis. The argument is based on two observations: first, that an interchange between b^i and b^{i+1} preserves the values of $F_j(b^j)$ for all indices other than i and i+1, and secondly, that for $F(x) = |x|$, the product $F_i(b^i)F_{i+1}(b^{i+1})$ is constant when the vectors b^i and b^{i+1} are exchanged. This permits us to deduce that $D(b^1,\ldots,b^n) = \Pi(F_i(b^i))^{n-i}$ decreases by a factor of $(1-\epsilon)$ at each interchange. It is easy to show that $\Pi(F_i(b^i))^{n-i} \geq 1$, for any basis, from which the polynomial convergence follows readily.

Constancy of $F_i(b^i)F_{i+1}(b^{i+1})$ is not valid for a general distance function, and the basis reduction algorithm is not known to execute in polynomial time in the number of variables n. But the algorithm may be shown to be polynomial in the data of the problem for <u>fixed n</u>. We present two arguments for this conclusion, both of which depend on establishing lower bounds for the possible values assumed by $F_i(b^i)$ during the course of the algorithm.

To obtain such a lower bound, assume that $C \subset B(R)$, the ball of radius R. Then $F(x) \geq |x|/R$. Now let $b^1,\ldots,b^n$ be any basis for the lattice satisfying $F_i(b^i) \leq U$, and let $c^1,\ldots,c^n$, with $c^{i+1} = b^{i+1} + \Sigma_{j=1}^{i}\mu_{i+1,j}b^j$, be the basis constructed in the proof of Theorem 3, which satisfies $F_i(c^i) = F_i(b^i)$ and $F_1(c^i) \leq F_i(b^i) + (1/2)\Sigma_1^{i-1}F_j(b^j) \leq nU$. We have, therefore, $|c^i| \leq nUR$.

We estimate $F_i(b^i)$, from below, as follows:

$$\begin{aligned} F_i(b^i) &= \min F(b^i + \alpha_1 b^1 + \ldots + \alpha_{i-1} b^{i-1}) \\ &= \min F(b^i + \alpha_1 c^1 + \ldots + \alpha_{i-1} c^{i-1}) \\ &\geq \min |(b^i + \alpha_1 c^1 + \ldots + \alpha_{i-1} c^{i-1})|/R. \end{aligned}$$

But $\min |(b^i + \alpha_1 c^1 + \ldots + \alpha_{i-1} c^{i-1})|$ is the distance between the vector b^i and the space $\langle c^1, \ldots c^{i-1} \rangle$ and is therefore equal to

$$[G(c^1, \ldots, c^{i-1}, b^i)/G(c^1, \ldots, c^{i-1})]^{1/2},$$

where the Grammian

$$G(x^1, \ldots, x^i) = \det[(x^j, x^k)]_{j,k=1}^i.$$

Since $c^1, \ldots, c^{i-1}$ and b^i are integral, $G(c^1, \ldots, c^{i-1}, b^i) \geq 1$. Moreover,

$$G(c^1, \ldots, c^{i-1})^{1/2} \leq |c^1| \ldots |c^{i-1}| \leq (nUR)^{i-1}.$$

It follows that

$$F_i(b^i) \geq 1/[R(nRU)^{i-1}] \geq 1/[R(nRU)^{n-1}] = V.$$

We have already shown that each component of

$$F_1(b^1), \ldots, F_i(b^i), F_{i+1}(b^{i+1}), \ldots, F_n(b^n)$$

is bounded above by $U = \max F_i(a^i)$ throughout the course of the algorithm. Moreover, the first term in the sequence to change at any iteration decreases by a factor of $(1 - \varepsilon)$. Our first argument for polynomial convergence is to observe that the maximal number of interchanges is therefore

$$[\log(U/V)/\log(1/(1-\varepsilon))]^n.$$

(Simply record the times at which the first two basis vectors b^1 and b^2 are interchanged. Between any consecutive pair of such times we are faced with an identical problem with n-1 variables.) Using our particular lower bound V we see that the number of interchanges of the basis reduction algorithm is bounded above by

(6) $$[n\log(nUR)/\log(1/(1-\epsilon))]^n.$$

The second argument for polynomiality, which achieves a different bound, depends on the observation that for a general distance function F(x), the product $F_i(b^i)F_{i+1}(b^{i+1})$ increases by a factor less than or equal to 2 after an interchange of b^i and b^{i+1}. The argument makes use of the following Theorem.

Theorem 5. Let S be a compact convex set in R^k, which is symmetric about the origin, and let x and y be two linearly independent vectors on the boundary of S. Define

$$d_x = \max\{\alpha \mid \alpha x + \beta y \in S \text{ for some } \beta\} \text{ and}$$
$$d_y = \max\{\beta \mid \alpha x + \beta y \in S \text{ for some } \alpha\}.$$

Then $1/2 \le d_x/d_y \le 2$.

Proof: If $d_x x + \beta y \in S$, then either $\beta \ge d_x - 1$, or $\beta \le 1 - d_x$. For if $0 \le \beta < d_x - 1$, x is a strict convex combination of 0, -y, $d_x x + \beta y$, and is therefore interior to S; if $1 - d_x < \beta \le 0$, x is a strict convex combination of 0, y, $d_x x + \beta y$ and is again interior to S. In the first case, $d_y \ge d_x - 1$. In the second case, since S is symmetric, the vector $-(d_x x + \beta y) \in S$ and again $d_y \ge d_x - 1$. It follows that

$$d_x/d_y \le 1 + 1/d_y \le 2,$$

since $d_y \ge 1$. The lower inequality follows from interchanging x and y. ⊗

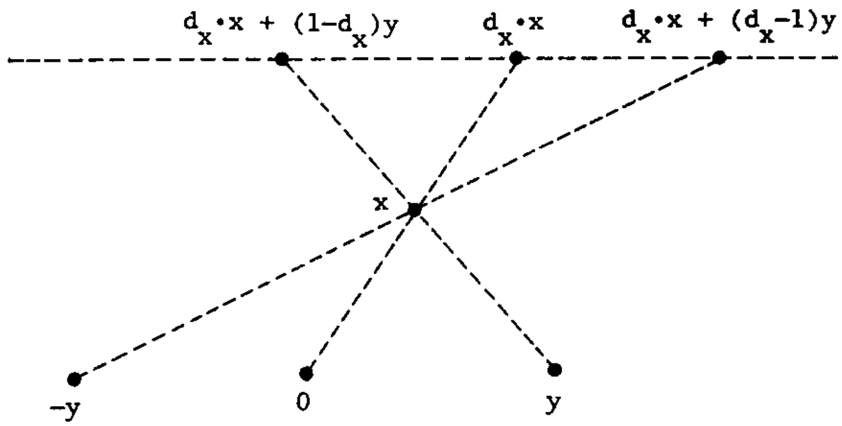

FIGURE 3

Theorem 5 may be used to show that the product

$$F_i(b^i)F_{i+1}(b^{i+1})$$

increases by a factor not larger than 2 at any step of the basis reduction algorithm in which b^i and b^{i+1} are interchanged. Let $S = C_i \subset E_i = \langle b^i,\ldots,b^n\rangle$. Assume, without loss of generality, that $F_i(b^i) = 1$, and take $y = b^i$ and $x = b^{i+1}/F_i(b^{i+1})$, both of which are on the boundary of C_i. But then $F_{i+1}(x) = 1/d_x$, and $F^{\#}_{i+1}(b^i) = 1/d_y$, with $F^{\#}_{i+1}$ the distance function associated with the projection of C into $\langle b^i, b^{i+2},\ldots,b^n\rangle$. It follows that

(7) $$F_i(b^{i+1})F^{\#}_{i+1}(b^i)/F_i(b^i)F_{i+1}(b^{i+1}) = [F_i(b^{i+1})/d_y]/[F_i(b^{i+1})/d_x] = d_x/d_y \le 2.$$

Now consider

$$D(b^1,\ldots,b^n) = \Pi(F_i(b^i))^{\gamma^{n-i}},$$

with $\gamma = 2 + 1/\log(1/(1-\varepsilon))$. It is a straightforward computation to show that our estimate (7) implies that $D(b^1,\ldots,b^n)$ decreases by a factor of at least $(1-\varepsilon)$ at each interchange required by the basis reduction algorithm.

Since $V \le F_i(b^i) \le U$ at each step of the algorithm, we see that the number of interchanges is bounded above by

$$[(\gamma^n-1)/(\gamma-1)]\log(U/V)/\log(1/(1-\varepsilon))$$
$$\le [(\gamma^n-1)/(\gamma-1)]n\log(nUR)/\log(1/(1-\varepsilon)),$$

an estimate which is much better than our previous estimate (6) in terms of its dependence on UR. The preceding discussion has established the following theorem:

Theorem 6. The basis reduction algorithm terminates in a polynomial number of steps, for fixed n.

Since the number of possible values of the vector

$$F_1(b^1),\dots,F_i(b^i),F_{i+1}(b^{i+1}),\dots,F_n(b^n)$$

is finite, the basis reduction algorithm executes in finite time even when $\varepsilon = 0$. Bárány has recently demonstrated geometric convergence when $\varepsilon = 0$ for the case of two variables. Consider two successive steps of the algorithm. Assume that the initial basis is (b^1,b^2), with b^2 the smallest lattice point on the line $b^2 + \alpha b^1$, and that $\delta_1 F(b^1) < F(b^2) < F(b^1)$. After the first interchange the basis is given by (b^2,b^1). Let μ^* minimize $F(b^1+\mu b^2)$ for integral μ and assume that $\delta_2 F(b^2) < F(b^1+\mu^* b^2) < F(b^2)$ so that another interchange is required leading to the basis $(b^1+\mu^* b^2,b^2)$. Finally, let μ be the integer which minimizes $F(b^2+\mu(b^1+\mu^* b^2))$.

Theorem 7 (Bárány). If $\delta_1\delta_2 > 1/2$ then $\mu = 0$ or 1. In either case the basis $(b^1+\mu^* b^2,b^2+\mu(b^1+\mu^* b^2))$ is reduced.

Proof: We argue, first of all, that $|\mu^*| > 1$. If $\mu^* = 0$, there is a contradiction between $F(b^1+\mu^* b^2) < F(b^2)$ and $F(b^2) < F(b^1)$. If $\mu^* = 1$, then b^2 is not the shortest integral vector on the line $b^2+\alpha b^1$, and similarly for $\mu^* = -1$. To be specific, let us now assume that $\mu^* \le -2$.

Consider the convex function $g(\alpha) = F(b^2+\alpha(b^1+\mu^* b^2))$. We have $g(0) = F(b^2)$ and from our assumptions $g(0) > \delta_1 F(b^1) > \delta_2\delta_1 F(b^1)$. Also $g(1) = F((b^1+(\mu^*+1)b^2) \geq F(b^1+\mu^* b^2) > \delta_2 F(b^2) > \delta_2\delta_1 F(b^1)$. But $g(-1/\mu^*) = (1/|\mu^*|)F(b^1) \leq (1/2)F(b^1)$. It follows, from the convexity of $g(\alpha)$, that if $\delta_1\delta_2 > 1/2$, the integral minimum of $g(\alpha)$ is at $\alpha = 0$ or $\alpha = 1$. In the first case, the basis $(b^1+\mu^* b^2, b^2)$ is reduced since $F(b^1+\mu^* b^2) < F(b^2)$; in the second case, the basis $(b^1+\mu^* b^2, b^1+(\mu^*+1)b^2)$ is reduced because $F(b^1+\mu^* b^2) \leq F(b^1+(\mu^*+1)b^2)$ ⊗

Theorem 7 implies that in 2p steps of the basis reduction algorithm, $F(b^1)$ will decrease by a factor of at least $(1/2)^p$. Since $F(h) \geq 1/R$ for any lattice point h, we have geometric convergence of the algorithm for $n = 2$ and $\varepsilon = 0$. No argument is currently available for higher dimensions, and $\varepsilon = 0$, unless we revise the order in which the steps of the algorithm are executed. For example, following a suggestion made by Bárány, let us assume that we always select the <u>largest</u> index i for which one of the conditions of a reduced basis is not satisfied. It follows that if we ever return to level 1, the basis $b^2,\ldots,b^n$ is reduced with $\varepsilon = 0$ for the n-1 dimensional problem defined by C_2. If an interchange of b^1 and b^2 is then required, two possible cases arise:

1. $F_1(b^2) \geq (1-\delta)F_1(b^1)$ for some fixed $1/2 < \delta < 1$. But then the basis $b^1,\ldots,b^n$ will be δ-reduced for the original problem. Our previous analysis shows that there are a finite number, $N(n,\delta)$ of lattice points h, such that $F_1(h) < F_1(b^1)$, and, therefore, the algorithm requires an exchange of b^1 and b^2 not more than $N(n,\delta)$ times.

2. At each return to level 1, we have $F_1(b^2) < (1-\delta)F_1(b^1)$, and

therefore the number of returns to level 1 is bounded above by

$$\log(U/V)/\log(1/(1-\delta)).$$

We then use an inductive argument on n to achieve polynomial bounds on the running time of the algorithm for $\varepsilon = 0$ and fixed n.

IV. The Korkine-Zolotarev Basis

A special basis for a lattice, the Korkine-Zolotarev basis, has been used very successfully by Lagarias, Lenstra and Schnorr (1990) to improve some classical estimates in the Geometry of Numbers relating the successive minima of a body C and its dual body $C^* = \{y \mid y.x \le 1 \text{ for all } x \in C\}$. In their analysis they approximate a general body by an ellipsoid, transform the ellipsoid to a sphere by a linear transformation and use specific properties of the spherical norm. We shall illustrate, by means of a few examples, that their arguments can be applied, virtually unchanged, to a general body without the prior step of an ellipsoidal approximation.

Let $b^1, b^2, \ldots, b^n$ be defined recursively as follows: given $b^1, \ldots b^{i-1}$, b^i minimizes $F_i(h)$ over all non-zero lattice points in $\langle b^i, \ldots, b^n \rangle$. The vectors $b^1, b^2, \ldots, b^n$ clearly form a basis, since otherwise there is an integer vector which can be written as a linear combination of the b^i with <u>some</u> fractional coefficients. But then by adding and subtracting suitable integral multiples of $\{b^j\}$, we obtain an integral vector

$$h = \alpha_1 b^1 + \ldots + \alpha_i b^i,$$

with α_i a proper fraction; $\alpha_i b^i$ is in the lattice projected into $\langle b^i, \ldots, b^n \rangle$ and gives a smaller value of $F_i(h)$ than does b^i.

Using this particular basis, the Korkine-Zolotarev basis is defined by

$$c^{i+1} = b^{i+1} + \Sigma_{j=1}^{i} \mu_{i+1,j} b^j,$$

as in (4). The basis satisfies the inequalities

(5) $$F_1(c^i) \le F_i(b^i) + (1/2)\Sigma_1^{i-1} F_j(b^j).$$

The Korkine-Zolotarev basis may not be unique; there may be several non-zero integral vectors in $<b^i,\ldots,b^n>$ which minimize $F_i(h)$, and the integers $\mu_{i,j}$ need not be uniquely defined.

Theorem 8. Let $c^1,\ldots,c^n$ be a Korkine-Zolotarev basis. Then

$$F_1(c^i)/((i+1)/2) \le \lambda_i \le ((i+1)/2)F_1(c^i).$$

Proof: Let $h^1,\ldots,h^n$ realize the successive minima. For each i, at least one of the vectors $h^1,\ldots,h^i$ must project to a non-zero lattice point in $<b^i,\ldots,b^n>$, since otherwise the vectors would all lie in $<b^1,\ldots,b^{i-1}>$ and be linearly dependent. It follows that $\max_{j\le i} F_i(h^j) \ge F_i(b^i)$ and therefore $\lambda_i = F_1(h^i) = \max_{j\le i} F_1(h^j) \ge F_i(b^i)$. But then (5) implies that $F_1(c^i) \le \lambda_i + (1/2)(\lambda_1 + \ldots + \lambda_{i-1}) \le ((i+1)/2)\lambda_i$. This demonstrates the left hand inequality of Theorem 8.

To obtain the right hand side, notice that for $k \le i$, $F_k(b^k) \le F_k(c^i) \le F_1(c^i)$, since c^i projects into a non-zero lattice point in $<b^k,\ldots,b^n>$. But

$$\begin{aligned}\lambda_i &\le \max_{j\le i} F_1(c^j)\\ &\le \max_{j\le i}\{F_j(b^j) + (1/2)\Sigma_{k\le j-1} F_k(b^k)\}\\ &\le F_1(c^i)\max_{j\le i}\{1 + (1/2)\Sigma_{k\le j-1} 1\}\\ &= ((i+1)/2)F_1(c^i) \quad \otimes\end{aligned}$$

We remark that Theorem 8, in conjunction with Minkowski's inequality, implies that a Korkine-Zolotarev basis satisfies

$$F_1(c^1).F_1(c^2)\ldots F_1(c^n).\mathrm{vol}(C) \le (n+1)!,$$

an improvement over the estimate of Theorem 4.

Let λ_1^* be the length of the shortest non-zero lattice point with respect to the dual body $C^* = \{y \mid y.x \le 1 \text{ for all } x \in C\}$. Minkowski's first theorem implies that $\lambda_1 \leq 2/(\mathrm{vol}(C))^{1/n}$ and $\lambda_1^* \leq 2/(\mathrm{vol}(C^*))^{1/n}$, so that an upper bound for $\lambda_1\lambda_1^*$ may be obtained from a lower bound for the product of the volumes $\mathrm{vol}(C).\mathrm{vol}(C^*)$. A well-known ellipsoidal approximation to C is sufficient to produce the inequality $\lambda_1\lambda_1^* \leq n^{3/2}$. A more sophisticated lower bound, quoted by Kannan and Lovász (1988), implies that there exists a universal constant c_0, such that $\lambda_1\lambda_1^* \leq c_0 n$. This result is used to demonstrate the following property of a Korkine-Zolotarev basis.

Theorem 9. There is a universal constant c_0 such that for a Korkine-Zolotarev basis, $F_i(b^i)\lambda_1^* \leq c_0(n-i+1)$.

Proof: We assume, without loss of generality, that the Korkine-Zolotarev basis consists of the n unit vectors $e^1,\ldots,e^n$, and let C_i be the projection of C into $e^i,\ldots,e^n$, with associated distance function F_i. The projection of the original lattice is the set of all $(x_i,\ldots,x_n)$ with integral coordinates. For this lattice and distance function, $\lambda_1 = F_i(b^i)$.

From the previous discussion, there is a non-zero lattice point $h^i = (h_i,\ldots,h_n)$ such that $F_i(b^i).\max\{h^i.x \mid x \in C_i\} \leq c_0(n-i+1)$. But this linear function $h^i.x$ may be extended to a linear function h.x in R^n by adding i-1 zero coordinates to h^i, so that

$$F_i(b^i)\lambda_1^* \leq F_i(b^i).\max\{h.x \mid x \in C\} \leq c_0(n-i+1). \quad \otimes$$

Theorem 9 has an important application to the study of lattice free bodies K which are not symmetric about the origin. As we shall see, any such body has associated with it a non-zero lattice point h such that

$$\max_{x \in K}\{h.x\} - \min_{x \in K}\{h.x\} \leq c_0 n(n+1)/2.$$

The argument is based on Theorem 9, and a subsequent result in the paper by Kannan and Lovász which may be described as follows.

Theorem 10. Let $C = (K - K)$, with K a convex body, and let F be the distance function associated with C. For any basis $b^1,\dots,b^n$, define $\rho = \Sigma F_i(b^i)$. Then the lattice translates of ρK cover R^n.

Proof: We show, by induction on n, that for any $x \in R^n$, there is a lattice point h with $x+h \in \rho K$. Notice that the hypotheses and conclusion of the Theorem are unchanged if we replace K by any translate of itself; we may therefore assume that K has been translated so that both 0 and b^1 are contained in $F_1(b^1)K$. Let K′ be the projection of K along the vector b^1 into $\langle b^2,\dots,b^n\rangle$ and x′ the corresponding projection of x.

By the induction assumption, there is a lattice point h′ in $\langle b^2,\dots,b^n\rangle$ such that $x'+h' \in \Sigma_2^n F_i(b^i)K'$ and therefore $x+\alpha b^1+h' \in \Sigma_2^n F_i(b^i)K$ for some α. It follows that $x+\lceil\alpha\rceil b^1+h' \in \Sigma_2^n F_i(b^i)K + (\lceil\alpha\rceil-\alpha)b^1 \subseteq \Sigma F_i(b^i)K$. ⊗

If the body K is free of lattice points, then its lattice translates do not cover the origin, and therefore $\rho = \Sigma F_i(b^i) > 1$. We see from Theorem 9, that for such a body, $\lambda_1^* < c_0\Sigma(n-i+1) = c_0 n(n+1)/2$. It should be remarked that the inductive argument provides an algorithm for calculating a lattice point in K, if $\rho \le 1$.

REFERENCES

Bárány, I., private communication.

Kannan, R. and L. Lovász, Covering Minima and Lattice-Point-Free Convex Bodies, Annals of Mathematics, 128, pp. 577-602 (1988).

Lagarias, J., H. W. Lenstra, Jr. and C. P. Schnorr, Korkine-Zolotarev Bases and Successive Minima of a Lattice and its Reciprocal Lattice, manuscript (1986).

Lenstra, H. W., Jr., Integer Programming with a Fixed Number of Variables, Mathematics of Operations Research, 8, pp. 538–548 (1983).

Lenstra, A. K., H. W. Lenstra, Jr. and L. Lovász, Factoring Polynomials with Rational Coefficients, Mathematische Annalen, 261, pp. 513–534 (1982).

Lovász, L., Geometry of Numbers and Integer Programming, Manuscript prepared for the 13th Annual Symposium on Mathematical Programming, 1988.